Highly Heterogeneous Rates of Evolution in the *SKP1* Gene Family in Plants and Animals: Functional and Evolutionary Implications

Hongzhi Kong,†‡§ Jim Leebens-Mack,†‡§ Weimin Ni,†‡ Claude W. dePamphilis,†‡§ and Hong Ma†‡§

*Laboratory of Systematic and Evolutionary Botany, Institute of Botany, The Chinese Academy of Sciences, Beijing, China; †Department of Biology, ‡The Huck Institute for Life Sciences; and §Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park, Pennsylvania

Skp1 (**S**-phase **k**inase-associated **p**rotein **1**) is a core component of SCF ubiquitin ligases and mediates protein degradation, thereby regulating eukaryotic fundamental processes such as cell cycle progression, transcriptional regulation, and signal transduction. Among the four components of the SCF complexes, Rbx1 and Cullin form a core catalytic complex, an F-box protein acts as a receptor for target proteins, and Skp1 is an adaptor between one of the variable F-box proteins and Cullin. Whereas protists, fungi, and some vertebrates have a single *SKP1* gene, many animal and plant species possess multiple *SKP1* homologs. It has been shown that the same Skp1 homolog can interact with two or more F-box proteins, and different Skp1 homologs from the same species sometimes can interact with the same F-box protein. In this paper, we demonstrate that multiple Skp1 homologs from the same species have evolved at highly heterogeneous rates. Parametric bootstrap analyses suggested that the differences in evolutionary rate are so large that true phylogenies were not recoverable from the full data set. Only when the original data set were partitioned into sets of genes with slow, medium, and rapid rates of evolution and analyzed separately, better-resolved relationships were observed. The slowly evolving Skp1 homologs, which are relatively highly conserved in sequence and expressed widely and/or at high levels, usually have very low d_N/d_S values, suggesting that they have evolved under functional constraint and serve the most fundamental function(s). On the other hand, the rapidly evolving members are structurally more diverse and usually have limited expression patterns and higher d_N/d_S values, suggesting that they may have evolved under relaxed or altered constraint, or even under positive selection. Some rapidly evolving members may have lost their original function(s) and/or acquired new function(s) or become pseudogenes, as suggested by their expression patterns, d_N/d_S values, and amino acid changes at key positions. In addition, our analyses revealed several monophyletic groups within the *SKP1* gene family, one for each of protists, fungi, animals, and plants, as well as nematodes, arthropods, and angiosperms, suggesting that the extant *SKP1* genes within each of these eukaryote groups shared only one common ancestor.

Introduction

In eukaryotes, the ubiquitin-mediated proteolysis by the 26S proteosome is one of the most fundamental regulatory mechanisms controlling cell cycle progression, transcriptional regulation, signal transduction, and many other cellular processes (Hershko and Ciechanover 1998; Callis and Vierstra 2000; Zheng et al. 2002). The ubiquitination of targets for degradation is a three-enzyme process, in which the E1 ubiquitin-activating enzyme and E2 ubiquitin-conjugating enzyme are relatively nonspecific, whereas different E3 ubiquitin ligases recognize specific substrates (Koepp et al. 2001). The largest known group of E3 ubiquitin ligases is called the SCF (**S**kp1, **C**ullin, and an **F**-box protein) complexes, which also contain a fourth subunit, Rbx1 (Bai et al. 1996; Willems et al. 1999; Zheng et al. 2002). Rbx1 and Cullin form a core catalytic complex, an F-box protein acts as a factor for recognizing target proteins, and Skp1 (S-phase kinase-associated protein 1) serves to link one of the variable F-box proteins to Cullin (Bai et al. 1996; Krek 1998; Zheng et al. 2002).

There is only one known functional Skp1 protein in human and yeasts. Nevertheless, this unique protein is able to interact with different F-box proteins to ubiquinate different substrates (Schulman et al. 2000; Ilyin et al. 2002). The human Skp1, for example, was originally identified as a protein that associates the cyclin A-CDK2 (cyclin-dependent kinase 2) complex with the F-box protein Skp2, and the SCF^{skp2} complex can ubiquitinate the $p27^{Kip1}$ inhibitor of CDK, allowing CDK activity to drive cells into S phase (Ganoth et al. 2001; Koepp et al. 2001). Another well-studied SCF complex in human, SCF^{FWD1}, targets the NF-κB pathway and β-catenin, a downstream signaling factor in the Wnt pathway of development and proliferation (Hatakeyama et al. 1999; Hattori et al. 1999). In budding yeast, *SKP1* is essential for cell cycle progress; SCF^{cdc4} has been shown to be involved in the degradation of the Cdk inhibitor Sic1 and transcriptional regulator Gcn4 (Perkins, Drury, and Diffley 2001).

Unlike human and yeasts, *Arabidopsis thaliana*, *Caenorhabditis elegans*, and *Drosophila melanogaster* have multiple *SKP1* homologs (Farras et al. 2001; Nayak et al. 2002; Yamanaka et al. 2002). In *A. thaliana*, the *SKP1* homolog *ASK1* (***A**rabidopsis **SK**P1-LIKE **1***) gene is part of the SCF^{TIR1} and SCF^{COI1} complexes that regulate the responses to plant hormones auxin and jasmonate, respectively (Gray et al. 1999; Xu et al. 2002). Genetic studies also revealed that *ASK1* is important for vegetative and flower development and essential for male meiosis (Samach et al. 1999; Yang et al. 1999; Zhao et al. 1999). In addition to *ASK1*, the *Arabidopsis* genome has at least 18 other *ASK* genes, and recent studies indicated that some of these Ask proteins have quite different ability to interact

Key words: *SKP1* gene family, rate heterogeneity, evolution, *Arabidopsis*.

E-mail: hxm16@psu.edu.

Mol. Biol. Evol. 21(1):117–128. 2004
DOI: 10.1093/molbev/msh001
Molecular Biology and Evolution vol. 21 no. 1

with F-box proteins (Gagne et al. 2002). Similarly, seven and 21 *SKP1* homologs have been predicted in the *D. melanogaster* and *C. elegans* genomes, respectively. Two of the *C. elegans* ***SK****P1-****R****ELATED* genes, *SKR-1* and *SKR-2*, play essential roles in both mitosis and meiosis, whereas *SKR-7*, *SKR-8*, *SKR-9*, and *SKR-10* are required for other aspects of normal development (Nayak et al. 2002; Yamanaka et al. 2002).

The fact that protists and fungi have a single *SKP1* gene, whereas many animal and plant species possess multiple members, suggests that the ancestral *SKP1* has duplicated, giving rise to diverse gene families in various lineages of multicellular organisms. Recently, a preliminary study on the phylogeny of 30 *SKP1* homologs from five fungal and animal species led to the conclusion that all *C. elegans SKR* genes form a monophyletic group and all the *D. melanogaster SKP1* homologs form a single clade (Nayak et al. 2002). If this is the case, then gene duplication occurred independently in these two animal lineages. This led us to ask whether the *SKP1* homologs from plant species, such as *A. thaliana* or rice, comprise distinct lineage-specific clades. Because the *SKP1* homologs from fungi, animals, and plants play very important roles in such fundamental processes as mitosis and meiosis, did duplicate genes maintain ancestral function over evolutionary time? Alternatively, have some genes lost their original function and/or acquired new functions after duplication? Given the key roles of *SKP1* homologs in transcriptional regulation, signal transduction, cell cycle processes, improved understanding of the evolutionary history of this gene family will provide insights into the origin and diversification of multicellular organisms.

Recently, the genomes of two rice subspecies have been sequenced (Goff et al. 2002; Yu et al. 2002), providing an excellent opportunity to compare the entire sets of *SKP1* homologs of rice and *Arabidopsis* as well as fungi and animals. In this paper, we report new results from studies of structure, phylogeny, and expression of *SKP1* genes. Our analyses indicate that, at least in animal and plant lineages, the multiple Skp1 homologs from the same species have evolved at highly heterogeneous rates. The slowly evolving members are relatively conserved in sequence and are expressed widely and/or at high levels, suggesting that they may have the most fundamental function(s). On the other hand, moderately and rapidly evolving members are structurally much more diverse and are usually more limited in expression, suggesting that they may have lost their original function(s) and/or acquired new function(s).

Materials and Methods

Data Retrieval

A search for the Skp1 homologs was performed using BlastP, tBlastN, and PsiBlast programs (Altschul et al. 1997; Schaffer et al. 2001) at the NCBI Web site against the nonredundant (nr), expressed sequence tag (EST), and high throughput genome sequence (htgs) databases on December 20, 2002. We started our search with the yeast and human Skp1 proteins and then continued with various published Skp1 homologs from *C. elegans*, *D. melanogaster*, and *A. thaliana* to identify as many as possible Skp1 homologs from protists, fungi, animals, and plants. In particular, searches were conducted for putative or predicted Skp1 homologs from those species whose genome sequences are available. Only full-length or nearly full-length protein sequences were selected for analysis, with an E-value cutoff of 1e-10.

Sequence Alignment

Protein sequences were aligned with ClustalX version 1.81, using BLOSUM 30 as the protein weight matrix (Henikoff and Henikoff 1992). Several values for both gap opening and gap extension penalties were tried and compared to identify the commonly resolved domains. A combination of 9.0 and 0.3 was finally adopted because conserved domains were aligned most accurately with few gaps. Minor corrections were then performed manually on the alignment based on the crystal structure of the human Skp1 (Schulman et al. 2000; Zheng et al. 2002). Several regions were excluded because of ambiguous alignment, leaving an E-value total of 155 amino acids residues in our final analyses.

Phylogenetic Analyses

Phylogenetic analyses of the *SKP1* gene family based on amino acid sequences were carried out using neighbor-joining (NJ), maximum-parsimony (MP), and maximum-likelihood (ML) methods in MEGA version 2.1 (Kumar et al. 2001), PAUP* version 4.10b (Swofford 2001), and PHYLIP version 3.6a3 (Felsenstein 2002), respectively. NJ analyses were done with the ''pairwise deletion'' or ''complete deletion'' option selected, and distance estimates were calculated using a poisson correction or a variable rates correction with gamma set at 1.0. For MP analyses, 10,000 replicates of random stepwise addition with SPR (subtree pruning-regrafting) branch-swamping were performed by using Heuristic Search, with no more than two trees saved at each replicate, followed by TBR (tree bisection reconnection) branch-swapping and saving all trees, with all trees in memory as starting point. For ML analysis, trees were estimated by the proml program within PHYLIP using the Dayhoff substitution model (Dayhoff, Schwartz, and Orcutt 1978) with and without a correction for variable rates across sites and with gamma set at 1.0. Support for each node was tested with bootstrap (BS) analysis, 1,000 replicates for NJ, 250 for MP, and 100 for ML, using random input order for each replicate.

An ML phylogeny of *ASK* genes was also estimated based on both amino acid and nucleotide sequences, with Skp1 homologs from pine and the fern *Ceratopteris* as outgroups. The nucleotide-based analysis was done in PAUP* using the HKY+gamma model with a ML estimate of 1.0 for gamma, and the amino acid analysis was performed using proml as described above. Bootstrap support was estimated based on 100 replicates for both data sets.

Rate Evaluation

Considering that rate heterogeneity and long-branch attraction may have contributed to uncertainty in phylogeny

estimates and disagreement among methods, we applied a likelihood ratio test for significant variation in substitution rates across the ML phylogeny. The ML topology estimated in proml was loaded as a user-specified tree in promlk, and the difference of the log likelihoods as estimated with and without an enforced clock was doubled and compared with the χ^2 distribution with $n - 2$ (150) degrees of freedom (Felsenstein 2002). Sources of significant rate heterogeneity were localized in the phylogenies by applying the two-cluster (TCT) and branch-length tests (BLT) in LINTREE (Takezaki, Rzhetsky, and Nei 1995). The two-cluster test examines the hypothesis of the molecular clock for the two lineages above each interior node of a tree, whereas the branch-length test examines the deviation of each root-to-tip branch length relative to the average length (Takezaki, Rzhetsky, and Nei 1995). The two-cluster and branch-length tests were applied using both poisson and gamma distance corrections.

Parametric bootstrap analyses were performed to assess whether the considerable rate heterogeneity observed among lineages might have confounded phylogenetic analyses susceptible to long-branch attraction (Huelsenbeck, Hillis, and Jones 1996). The Pseq-Gen simulation program (Grassly, Adachi, and Rambaut 1997) was used to generate 100 data sets simulated on the ML phylogeny (topology and branch lengths), assuming the substitution probability matrix of Dayhoff, Schwartz, and Orcutt (1978). Each simulated data set was analyzed using the protdist and neighbor, proml and protpars programs within PHYLIP. Distances and ML phylogenies were estimated under the Dayhoff substitution model that was used to generate the sequences. A majority rule consensus tree was constructed and inspected for the frequency with which the analysis was able to recover the monophyletic clades observed in the ML phylogeny and expected for nematodes, arthropods, and angiosperms.

To further elucidate the evolutionary fates of duplicate *SKP1* homologs, ratios of nonsynonymous to synonymous nucleotide substitutions ($\omega = d_N/d_S$) were estimated for *ASK* genes using the codeml program of PAML version 3.13 (Yang 1997). A likelihood ratio test was used to test whether *ASK* genes on short branches identified in the LINTREE analysis were significantly more constrained than other *ASK* genes.

Expression Studies

Plants of *A. thaliana* (Columbia ecotype) were used to examine expression patterns of *ASK* genes. RNA was isolated from 3-day-old to 4-day-old seedlings, roots of 2-week-old plants, leaves of 3-week-old to 4-week-old plants, inflorescence stems and inflorescences of 5-week-old plants, and 2-day to 10-day postpollination siliques. Total RNA was isolated using the RNeasy mini kit (QIAGEN, Valencia, Calif.) and were treated with DNase I (Life Technologies/Gibco-BRL, Carlsbad, Calif.). The first-strand cDNA was synthesized using Super Script II reverse transcriptase (Life Technologies/Gibco-BRL) and used as a template for PCR with gene-specific primers (table 1 in Supplementary Material online). The specificity of the primers was verified by sequencing the PCR products from genomic DNA. PCR was carried out under standard conditions using 10 pmol of each primer and 35 cycles of 94°C for 30 s, 56°C to 66°C for 40 s, and 72°C for 60 s. Thirty microliters of PCR products were separated on 1.5% agarose gels and visualized by UV. Adenine phosphoribosyl transferase 1 (*APT1*) gene was also amplified in all samples as a positive control. Negative control PCRs without reverse transcriptase did not produce any PCR bands. For most *ASK* genes, we used only one pair of primers, but for some genes, nested primers were utilized because the first run of PCR detected very weak band(s) or nothing.

Results

Skp1-like Proteins

A large collection of amino acid sequences was retrieved by Blast searches and then condensed to a set of 152 Skp1 homologs (table 2 in Supplementary Material online). No Skp1 homolog was found in Eubacteria or Archaea, and only a single Skp1 homolog was obtained from each of the two protists and 12 fungi. For *C. elegans*, *D. melanogaster*, human, and *A. thaliana*, we included the 21, 7, 1, and 19 previous published Skp1 homologs, respectively. One Skp1 homolog was also identified from the newly finished puffer fish (*Fugu rubripes*) genome, and 21 and 22 rice (***O****ryza sativa*) **Sk**p1 homologs (Osk) were obtained from the two subspecies, *indica* and *japonica*. Sequence comparison and phylogenetic analysis indicated that there are 25 distinct *OSK* genes in rice, including 11 identical between the two subspecies, seven similar, and seven unique to one or the other (fig. 1 in Supplementary Material online). Partial sequences were also detected in the genome of several animal and plant species, and a detailed analysis on these sequences will be published elsewhere.

Sequence alignments of the Skp1 homologs indicated that although some parts of the amino acid sequence are quite divergent, other regions are highly conserved (fig. 2 in Supplementary Material online). The conserved regions correspond to the three β-strands (S1 to S3) and eight α-helices (H1 to H8) in the structure of the human Skp1 in co-crystals with other SCF subunits (Schulman et al. 2000; Zheng et al. 2002). Many residues that were proposed to be important for interaction with Cullin and F-box proteins, such as the Asn 49, Leu 100, Ile 104, Leu 105, Asn 108, Tyr 109, Leu 116, Cys 120, Val 123, Ile 127, and Phe 139 of the human Skp1, are conserved in many of the Skp1 homologs from all major groups, suggesting that they may have similar structure and function to the human Skp1.

When multiple Skp1 homologs exist in a species, only one or few seem to be structurally more conserved than others. In *C. elegans*, for example, Skr-1 is 70% identical to the human Skp1 and contains all the conserved regions and residues, whereas other Skr proteins are less than 60% identical to the human Skp1 and occasionally lack some structurally conserved regions and residues. The Leu 105 of human Skp1 has been replaced by Val in Skr-7, Skr-8, Skr-9, Skr-10, and Skr-11; by Asn in Skr-15,

Skr-16, and Skr-20; and by Thr, Lys, Arg, and Glu in Skr-17, Skr-18, Skr-19, and Skr-21, respectively (table 1; fig. 2 in Supplementary Material online). Similarly, SkpA, Ask1/2, and Osk1 of *D. melanogaster*, *A. thaliana*, and rice, respectively, seem to be more conserved than any other Skp1 homologs of these species.

Phylogenetic Analyses

In the NJ tree estimated for the whole data set, four major groups were observed, although detailed relationships among and within these groups are not very well supported (fig. 3 in Supplementary Material online). Taxonomically, these four groups correspond to protists (Mycetozoa), fungi (Fungi), animals (Metazoa), and plants (Viridiplantae) (Baldauf et al. 2000).

Within the plant clade, Ask1 and Ask2 were clustered into a weakly supported group together with 26 short-branched Skp1 homologs from 15 other eudicot species, and Osk1 formed a fairly resolved group together with seven short-branched Skp1 homologs from five other monocot species; the remaining Ask and Osk proteins, however, have longer branches and formed several lineage-specific clusters. The gymnosperm, fern, bryophyte, and algal Skp1 homologs were unexpectedly scattered throughout the short-branch eudicot clade. Similarly perplexing patterns were found in the animal clade: short-branched Skr-1 and SkpA were clustered into separate arthropod and nematode groups, and long-branched Skp1 homologs from these two species tended to form species-specific clusters. The NJ trees estimated under the gamma correction model ($a = 1.0$) or with ''complete deletion'' option were not significantly different from the tree in fig. 3 in Supplementary Material online (data not shown).

In the most parsimonious trees, the monophyly of protist, fungal, and plant clades was also supported, whereas the animal Skp1 proteins were resolved as a paraphyletic group (fig. 4 in Supplementary Material online). However, we could not reject the monophyly of the animal group, because MP trees with animal Skp1 homologs constrained as monophyletic were only one step longer than the most-parsimonious trees (fig. 5 in Supplementary Material online). As in the NJ tree, short-branched members from closely related species usually clustered together, whereas long-branched members tended to form one or several lineage-specific groups. This pattern was also found in the ML trees estimated with and without a correction for variation in rates across sites (fig. 6 in Supplementary Material online).

Rate Heterogeneity and Long-Branch Attraction

All of our NJ, MP, and ML trees indicated a mixture of short and long branches within each major group, suggesting that members of the *SKP1* gene family may have evolved at quite different rates, and that long-branch attraction may have contributed to the uncertainty and disagreement among phylogenetic analyses. The possibility of a constant substitution rate across the phylogeny was soundly rejected in the likelihood ratio test ($\triangle = 7868.88$; $P < 0.001$). Moreover, the results of our parametric bootstrap analyses indicated that the observed rate heterogeneity was extreme enough to cause long-branch attraction. Simulations were performed assuming monophyletic clades of angiosperm, arthropod, and nematode Skp1 homologs (as suggested in the ML results), but these clades were only recovered for 6%, 11%, and 31% of our replicates in the NJ analysis, respectively. MP and ML analyses of the simulated data gave similar results: MP recovered angiosperm, arthropod, and nematode clades in 12%, 11%, and 27% of the analyses, whereas ML recovered these clades in 11%, 28%, and 35% of the analyses. Inspection of sample trees indicated that unrelated long branches were commonly grouped together in the incorrect trees, indicating clearly that long-branch attraction had overridden phylogenetic signal.

The two-cluster test applied in LINTREE showed that the paired clades above almost all the major interior nodes as well as some smaller nodes evolved at significantly heterogeneous rates (fig. 1). The branch-length test further indicated that 65 of 150 sequences (43.3%) evolved significantly slower than average, and that 32 of 150 sequences (21.3%) evolved significantly faster than average. Skp1 homologs sampled from the same animal or plant species usually evolved at quite different rates. In *C. elegans*, *D. melanogaster*, *A. thaliana*, and rice, Skr-1, Dm.SkpA, Ask1, Ask2, Ask4, and Osk1 evolved significantly slower than others, whereas Skr-6 to Skr-21, Dm.CG15800, and Osk2, Osk3, Osk7, Osk10, Osk11, Osk13, Osk14, Osk17 to Osk22, and Osk24 evolved significantly faster than others. The Skp1 homologs from fungal species, however, all evolved at moderate rates.

Given that 97 (or 70.3%) of the 138 Skp1 homologs from animals and plants were found to evolve at significantly faster or slower than average rates, and this variation was shown to be driving long-branch attraction, we partitioned the original data set into three subsets based on the results of the branch-length test. The first data set (S-BRANCH) included 65 genes that have significantly shorter than average root-to-top distances. The second (M-BRANCH) and the third (L-BRANCH) data sets, contained 53 moderately evolving Skp1 homologs and 32 genes with significantly longer than average root-to-tip distances, respectively. The two protist Skp1 genes from *Dictyostelium discoideum* and *Physarum polycephalum* were included as outgroup sequences in all three data sets. Amino acid positions corresponding to human Skp1 residues proposed to be important for interaction with Cullin and F-box proteins were found to be highly conserved in the S-BRANCH genes (fig. 2 and table 1). Members of M-BRANCH and L-BRANCH data sets, however, exhibited greater amino acid variation and reduced frequencies for the most common residues, and functionally different amino acids were often observed at the key positions as compared with the human Skp1.

Our analyses of three partitions led to better-resolved trees relative to the global phylogenies (fig. 3). In the NJ tree of the S-BRANCH data set, a plant lineage and an animal lineage are evident, and the bootstrap values for them are 100% and 88%, respectively (fig. 3*A*). Within the plant lineage, three Skp1 homologs from nonangiosperm

plants were resolved as a poorly supported (56%) cluster, whereas the remaining angiosperm genes formed a well-supported (93%) cluster, in which a eudicot group and a monocot group were also observed. Within the animal lineage, four sequences from nematodes formed a basal clade, followed by clades for echinoderms, arthropods, platyhelminthes, and vertebrates. In the NJ tree of the M-BRANCH data sets, three major lineages, corresponding to fungi, animals, and plants, respectively, were recognizable, and the fungal lineage possessed a closer affinity to animals than to plants (fig. 3*B*). Within the animal lineage, four and five moderately evolving Skp1 homologs from *C. elegans* and *D. melanogaster* formed a nematode and an arthropod clade, respectively. Within the plant lineage, a weakly supported angiosperm clade was recovered. The L-BRANCH data set contained all the significantly rapidly evolving members from animals and plants, and our analysis also obtained two major lineages, one for each of them (fig. 3*C*). The well-supported clades observed in all of the NJ trees were also found in the MP and ML phylogenies and bootstrap support values for all three analyses are shown in figure 3.

Because of rate heterogeneity and long-branch attraction, it has been very difficult to reconstruct the phylogeny of the *Arabidopsis ASK* genes using conventional NJ and MP methods. However, ML analyses based on protein and DNA matrixes resulted in very similar topologies (also similar to the NJ tree based on gamma distance). Since the topology of the bootstrap consensus trees for the nucleotide-based and amino acid-based analyses were identical (fig. 4), we used this topology as the constraint tree for the codeml analysis of changes in d_N/d_S. First we tested whether branches leading to Ask1, Ask2, and Ask4 indicated a significantly different level of constraint (d_N/d_S) than branches leading to all other Ask protiens. We rejected the null hypothesis of a single rate ratio for all Ask proteins in favor of the two rate ratio model ($\triangle = 147.75$; $P < 0.001$). In a follow-up analysis, separate rate ratios were estimated for each branch on the Ask phylogeny. Inspection of these rate ratios confirmed the inference from the two-rate ratio test: Ask1, Ask2, and Ask4 have evolved under strong purifying selection, whereas evolution of the remaining Ask genes has been less

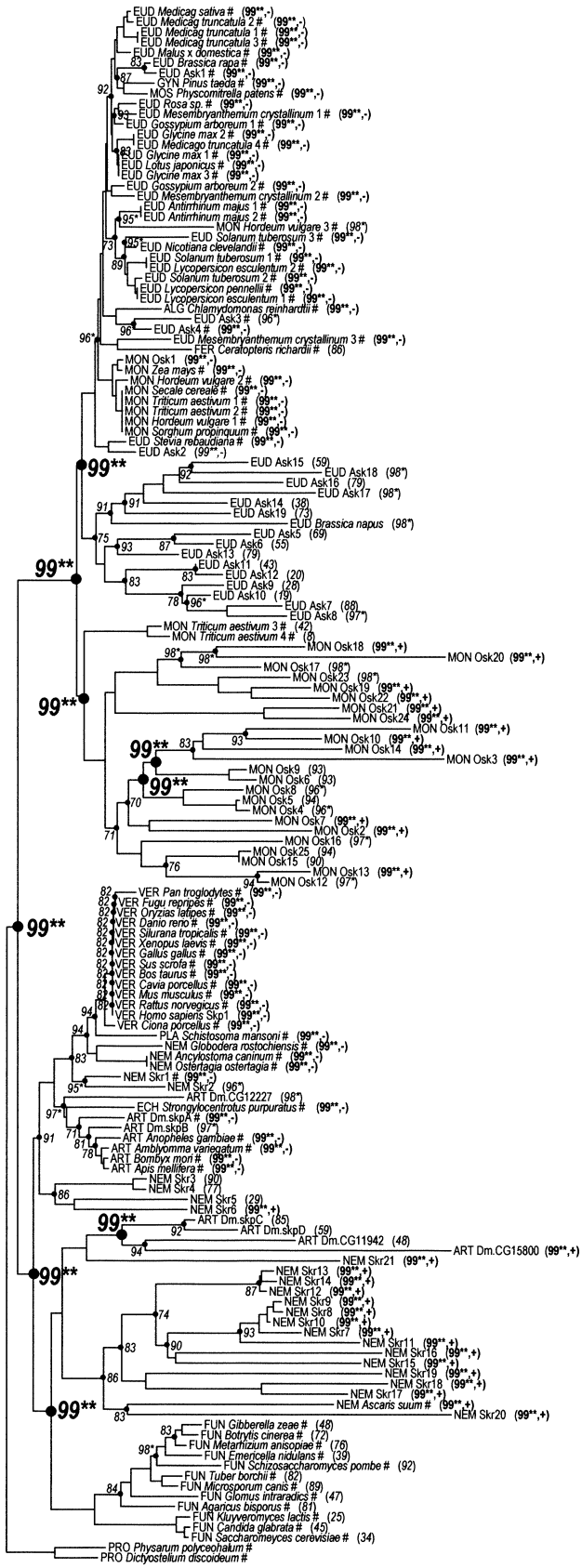

FIG. 1.—Highly heterogeneous rates of evolution of Skp1 homologs as indicated in the two-cluster test and branch-length tests of LINTREE. The two-cluster test showed that the paired clades above almost all the major interior nodes as well as some smaller nodes evolved at significantly heterogeneous rates. The branch-length test further indicated that 65 sequences evolved significantly slower than the average and that 32 sequences evolved significantly faster than the average. CP values for the two-cluster test higher than 70% are shown at the interior nodes concerned, and CP values for the branch-length test are shown in the parenthesis after each sequence name; ''99**'' means the significance level is 1%, and ''95*'' means the significance level is 5%. Plus (+) and minus (–) symbols indicated the sequences that evolved significantly faster and slower than the average, respectively, and ''#'' symbols indicate the sequences that have been detected in EST databases. The taxonomic categories for these organisms were abbreviated as follows: PRO = protists; FUN = fungi; NEM = nematodes; ART = arthropods; PLA = platyhelminthes; ECH = echinoderms; VER = vertebrates; ALG = green algae; MOS = mosses; FER = ferns; GYN = gymnosperms; EUD = eudicots; and MON = monocots.

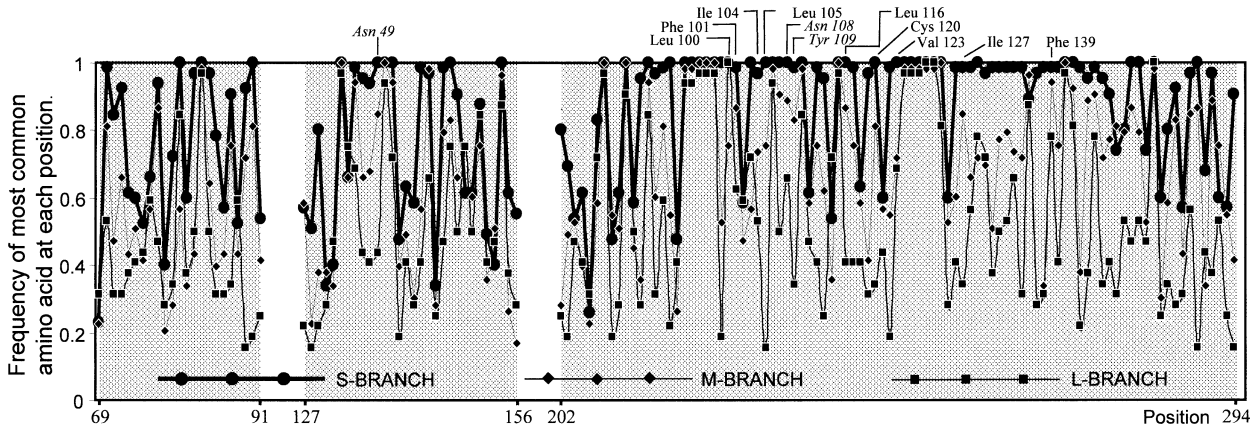

FIG. 2.—Comparison of S-BRANCH, M-BRANCH, and L-BRANCH data sets at the amino acid sequence level. For each data set, frequency of the most common amino acids at each position along the sequence was calculated. Information from two protists, *Dictyostelium discoideum* and *Physarum polycephalum*, was not included in the calculation. Only the relatively conserved regions used in phylogenetic analyses are shown, and six numbers along the X-axis indicate the N-terminal and C-terminal of the corresponding regions. The 12 residues that were proposed to be structurally important in interaction with Cullin and F-box proteins are marked in italic and regular font, respectively.

constrained. This inference is also supported by patterns of amino acid change at ''key'' positions of these *ASK* genes (fig. 4*B*). Structurally, Ask1 and Ask2 have the same amino acids as the human Skp1 at all the key positions that were proposed to be important for interaction with Cullin and F-box proteins, and Ask4 has only one change at Ile 127 (from I to M) (fig. 4*B*). The remaining Ask proteins exhibit greater amino acid variations and have reduced frequencies for the most common residues, sometimes with functionally different amino acids at the key positions (fig. 4*B*).

Expression Patterns

Because multiple Skp1 homologs from the same animal and/or plant species tend to evolve at quite heterogeneous rates, it is reasonable to expect that they exhibit different expression patterns and/or gene functions. Some paralogs may play the same fundamental role(s) as the single Skp1 protein in protists, fungi, and vertebrates, whereas others may have lost their original function(s) and/or acquired new function(s). To further understand the evolutionary fates of the Skp1 homologs, we first searched in the GenBank EST database for the expression information on the Skp1 homologs analyzed in the present study, then we investigated and compared the expression pattern of the 19 *ASK* genes from *A. thaliana*. Both studies indicated that, in some animal and plant species, when multiple Skp1 homologs exist, the rate heterogeneity of evolution was usually accompanied by the differentiation of expression pattern or gene function.

Our search of the GenBank EST database identified numerous ESTs that match exactly to one of 83 *SKP1* homologs. As shown in figure 1, 63 (96.9%) of the 65 significantly slowly evolving members have EST(s), whereas 89 (93.7%) of the 95 moderately and rapidly evolving members from animal and plant species do not. Nevertheless, the finding of one EST for a rapidly evolving *SKP1* homolog in the nematode species *Ascaris suum* suggested that some moderately and rapidly evolving members are still transcribed or even functional, but they are generally not expressed widely or at high levels.

In addition, our RT-PCR studies indicated that 19 *ASK* genes have distinct expression patterns (fig. 4*A*). *ASK1* and *ASK2* are expressed in all tissues tested and at relatively high levels. *ASK4*, *ASK11*, *ASK12*, and *ASK18* are also expressed widely but at lower levels. The expression of most other *ASK* genes can be detected in one or a few different tissues, but their patterns and/or levels are quite diverse. The expression of *ASK7*, *ASK8*, *ASK10*, *ASK16*, and *ASK17* is mainly in the silique, whereas that of *ASK5* is in the inflorescence. *ASK3* and *ASK14* are expressed in both inflorescence and silique, but *ASK3* has a higher expression level in silique, whereas *ASK14* has a higher level in inflorescence. *ASK9* and *ASK13* are similarly expressed in stem, inflorescence, and silique, although *ASK13* is also expressed in leaf. For *ASK11*, *ASK12*, and, *ASK15*, the first run of PCR detected very weak band(s) or nothing, although clear signals were finally observed when nested primers were utilized. No detectable RT-PCR product was found for *ASK6*, the shortest *ASK* gene that lacks the last two of eight α-helices (H7 and H8) (fig. 2 in Supplementary Material online), suggesting that it is expressed at very low levels if at all.

Discussion

Highly Heterogeneous Rate of Evolution

We have shown that members of the *SKP1* gene family, especially multiple *SKP1* homologs from the same animal or plant species, have evolved at highly heterogeneous rates. Such wide ranges of rate heterogeneity were evident in the global NJ, MP, and ML analyses (figs. 3–6 in Supplementary Material online) and verified by LINTREE tests (fig. 1). The two-cluster test indicated that significant rate heterogeneity exists at almost all the major interior nodes, and the branch-length test suggested that 97 (70.3%) of the 138 Skp1 homologs from animals and plants have evolved significantly slower or faster than average.

The high heterogeneity in rate of evolution confounded efforts to reconstruct the phylogenetic relationships among all identified members of the *SKP1* gene

family. In particular, long-branch attraction has contributed to the uncertainty in phylogenies and disagreement among phylogenetic analyses. The NJ, MP, and ML analyses based on all of the Skp1 homologs yielded quite different phylogenetic trees and, given the effects of long-branch attraction, it is very unlikely that any of them reflects the true history of the *SKP1* gene family. *SKP1* homologs on long branches tended to cluster together within plant and animal clades, and parametric bootstrap analyses further demonstrated that the observed variation in branch lengths significantly reduced the ability to reconstruct relationships specified in the simulated phylogenies. When members with similar evolutionary rates were analyzed separately, more highly supported ML, MP, and NJ phylogenetic trees were produced, and there was little conflict among the ML, MP, and NJ phylogenies. Data partitioning has reduced the effects of long-branch attraction.

Phylogenetic Relationships of the Skp1 Homologs

Phylogenetic analyses of three smaller data sets confirmed that the protist, fungal, animal, and plant Skp1 homologs each form separate monophyletic groups, suggesting that four major groups indicated in the global analyses may indeed reflect the evolutionary history of the *SKP1* gene family at very basal levels. If so, then the ancestors of eukaryotes as well as of the protist, fungal, animal, and plant lineages each shared as few as one *SKP1* gene. The protist, fungal, and vertebrate species included in this paper consistently have one copy of the *SKP1* gene, suggesting that the size of the gene family did not increase in these groups. The conservation of a single *SKP1* gene through long periods of evolutionary history suggests that either (1) these lineages did not experience gene duplications, or (2) all but one *SKP1* homologs have been lost in vertebrates, fungi, and the sampled protists. The phylogenetic position of distinct *SKP1* homologs sampled from many animal and plant species, however, clearly indicates that gene duplication occurred frequently and independently during the evolutionary history of these two lineages (fig. 3). Despite that, it is not yet clear whether the striking contrast between the few-gene and many-gene lineages is caused mainly by differences in gene birth rates, by gene death rates, or by both.

In NJ, MP, and ML analyses of partition data sets, distinct nematode and arthropod clades are strongly supported (fig. 3*A* and *B*), suggesting that multiple Skp1 homologs within these two lineages shared a single common ancestor and that gene duplication in these two lineages occurred separately. Similarly, it is very possible that the angiosperm Skp1 homologs shared a single common ancestor.

Within the angiosperm lineage, Skp1 homologs from eudicots and monocots formed two separate clades in S-BRANCH analysis (fig. 3*A*) but not in M-BRANCH analysis (fig. 3*B*), suggesting that the evolutionary history of angiosperm Skp1 homologs is still not clear. However, because the main relationships in figure 3*B* were not strongly supported by NJ, MP, or ML analysis, our results

Table 1
Amino Acid Conservation and Divergence Among Skp1 Homologs at 12 Critical Residues As Compared with the Human Skp1

Human		Asn 49	Leu 100	Phe 101	Ile 104	Leu 105	Asn 108	Tyr 109	Leu 116	Cys 120	Val 123	Ile 127	Phe 139
		N	L	F	I	L	N	Y	L	C	V	I	F
Protists[a]		N	L	F	I	L	N	Y	L	C	V	I	F
Fungi[b]		N	L	F/Y	I/V	L	N	Y	L	C	V	I	F
Caenorhabditis elegans	Skr-1	N	L	F	I	L	N	Y	L	C	V	I	F
	Others	N/K/E/R/T	L	F/V/T	I/V/A	L/V/N/E/K/R/T	N/R/H/T/L/F	Y/F/E/D/H/E/R	L/I/F/M	C/P/T	V/I/C/A	I/L/F/P/M/A	F/Y/L/E
Drosophila melanogaster	SkpA	N	L	F	I	L	N	Y	L	C	V	I	F
	Others	N/G/S/K	L/I	F/I/C	I/M/T	L/E	N/K	Y/Q	L/V/Y/E	C/Y	V/L	I/L	F
Other vertebrates[c]		N	L	F	I	L	N	Y	L	C	V	I	F
Arabidopsis thaliana	Ask1,2,4	N	L	F	I	L	N	Y	L	C	V	I/M	F
	Others	N/E/K	L/I/V	F/M	I/L/M/T	L/M/R/K/H/N/D	N/D/Y	Y/F/D	L/F/K	C/Y/F/A/G/S	V/I/C	I/L/V/M/T	F/L
Oryza sativa[d]	Osk1	N	L	F	I	L	N	Y	L	C	V	I	F
	Others	N/G/T/D/H/K	L/V	F/V/I/Y/H	I/L/T	L/M/N/R/T/A	N/D/E/T/A/S/H	Y/F/L/A/R	L/F/M/A	C/Y/V/A/S	V/I/L/A/T	I/L/M/T/S	F/L

[a] Including *Dictyostelium discoideum* and *Physarum polycephalum*.
[b] Including *Agaricus bisporus*, *Glomus intraradices*, *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Candida glabrata*, *Schizosaccharomyces pombe*, *Metarhizium anisopliae*, *Gibberella zeae*, *Botrytis cinerea*, *Microsporum canis*, *Tuber borchii*, and *Emericella nidulans*.
[c] Including *Ciona intestinalis*, *Oryzias latipes*, *Fugu rubripes*, *Danio rerio*, *Silurana tropicalis*, *Xenopus laevis*, *Gallus gallus*, *Sus scrofa*, *Bos taurus*, *Cavia porcellus*, *Mus musculus*, *Rattus norvegicus*, and *Pan troglodytes*.
[d] Including subspecies *indica* and *japonica*.

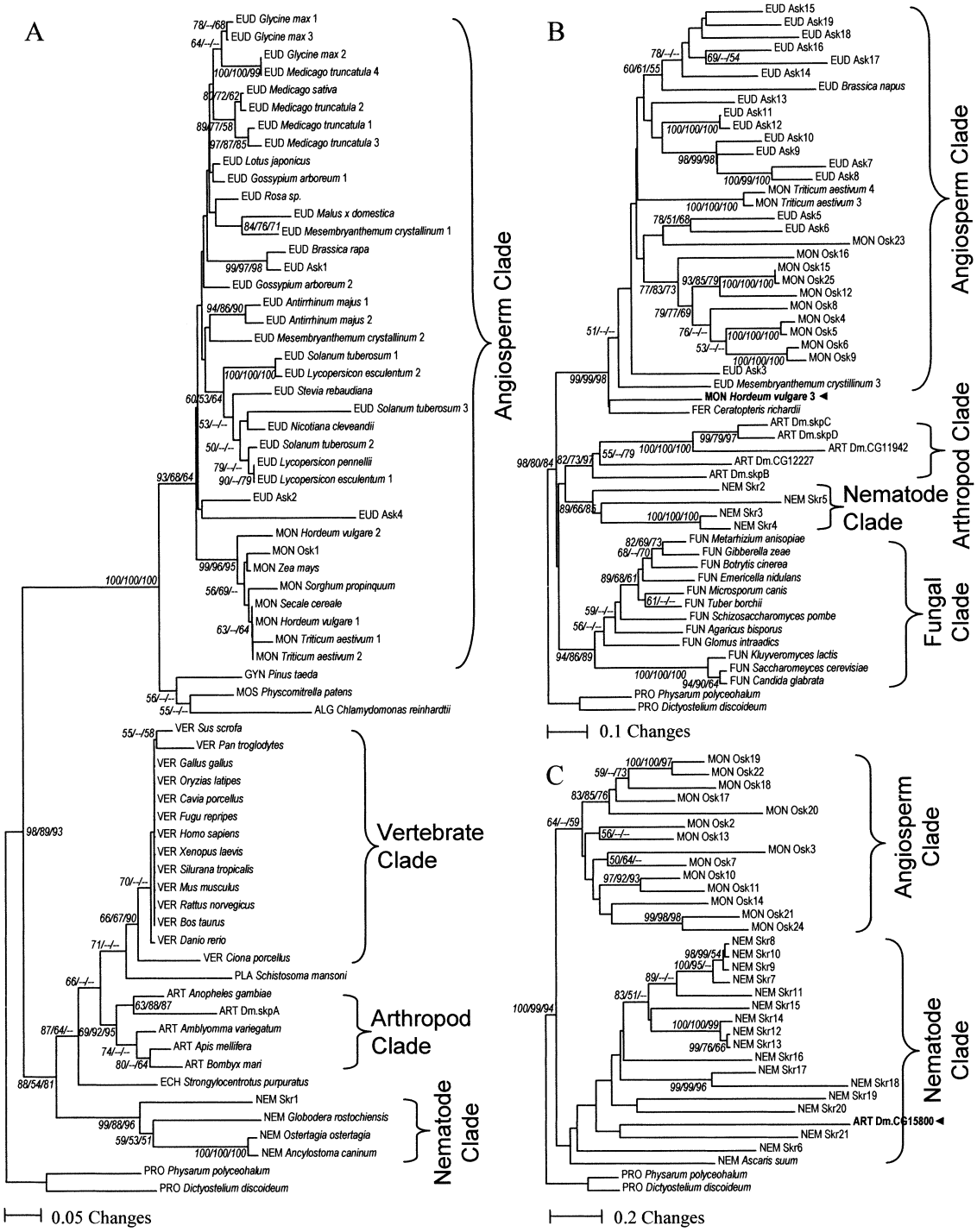

FIG. 3.—NJ trees for (*A*) S-BRANCH, (*B*) M-BRANCH, and (*C*) L-BRANCH data sets, with Bootstrap values (NJ/MP/ML) higher than 50% shown for each clade. The taxonomic categories for these organisms were abbreviated the same as in figure 1.

cannot reject the hypothesis that the eudicot and monocot Skp1 homologs will each form a single clade. Nevertheless, a cluster of short-branched Skp1 homologs from such distantly related groups as the Caryophyllales (*Mesembryanthemum*), Rosales (*Rosa* and *Malus*), Malvales (*Gossypium*), Fabales (*Glycine*, *Lotus*, and *Medicago*), Solanales (*Solanum*, *Lycopersicon*, and *Nicotiana*), Lamiales (*Antirrhinum*), and Asterales (*Stevia*) (fig. 3*A*), suggests a possible rapid diversification of Skp1 genes at the time or shortly after the origin of eudicots.

Functional and Evolutionary Implications

The most important known function of Skp1 and its closest homologs is its role in the SCF ubiquitin ligases; the Skp1 protein serves as an adaptor between Cullin and an F-

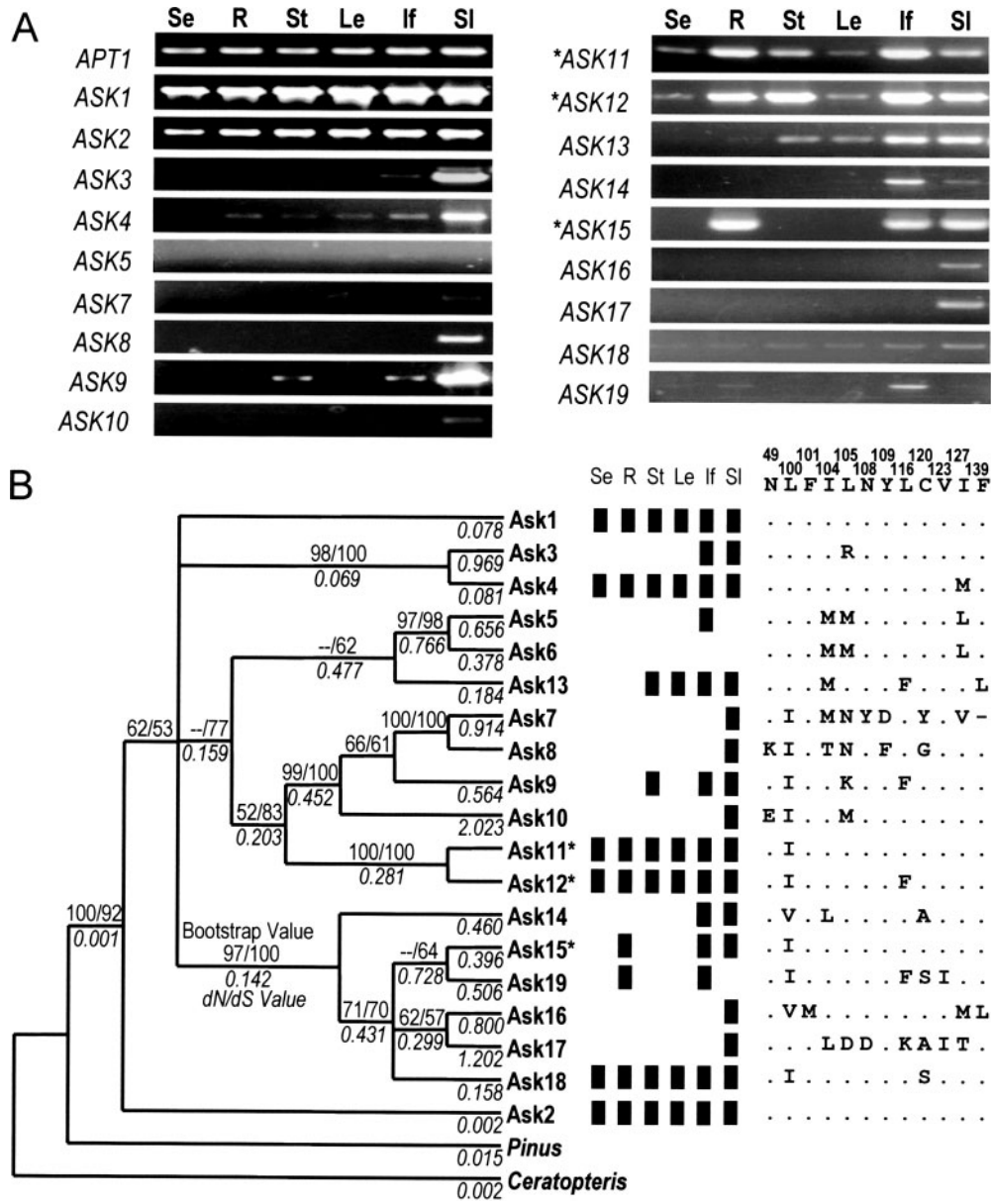

FIG. 4.—Expression and evolution of *ASK* genes. (*A*) Expression pattern of *ASK* genes as revealed by RT-PCR experiments, with adenine phosphoribosyl transferase 1 *(APT1)* gene as control. The plant organs are seedling (Se), root (R), stem (St), leaf (Le), inflorescence (If), and silique (Sl). For most *ASK* genes, we used only one pair of primers, but for *ASK11, ASK12*, and *ASK15* (as indicated by asterisks), nested primers were utilized because the first run of PCR detected very weak band(s) or nothing. (*B*) Phylogenetic relationships, rate heterogeneity (d_N/d_S values), expression pattern, and amino acid changes at the key positions of *ASK* genes.

box protein, allowing the enzyme to interact with its substrate(s) (Jackson and Eldridge 2002; Zheng et al. 2002). Previous studies indicated that the number of the F-box proteins is usually larger than that of Skp1 homologs (Kipreos and Pagano 2000; Gagne et al. 2002; Ilyin et al. 2002). In yeasts, at least three of the 14 F-box proteins have been found to function in the SCF complexes, whereas in human, four of the 38 F-box proteins have been found to interact with the Skp1 protein (Kipreos and Pagano 2000). Furthermore, 24 and 337 F-box genes have been identified in *D. melanogaster* and *C. elegans* genome, respectively (Kipreos and Pagano 2000). In *Arabidopsis*, although there are only 19 *SKP1* homologs, about 700 F-box genes have been predicted; furthermore, several *Arabidopsis* F-box proteins have been shown to interact with the Ask1 protein (Gray et al. 1999; Samach et al. 1999; Zhao et al. 1999; Zhao et al. 2001; Gagne et al. 2002; Xu et al. 2002). Therefore, in fungi, animals, and plants, one Skp1 protein can interact with several different F-box proteins. More recent studies also indicate that different Skp1 homologs from the same species sometimes can interact with the same F-box protein (Gagne et al. 2002).

As mentioned above, previous studies on the crystal structure of a human SCF complex suggested that some of the Skp1 residues are important for interaction with the Cullin and F-box proteins (Schulman et al. 2000; Zheng

et al. 2002). From figure 2 and from figure 2 in Supplementary Material online, it is evident that these conserved residues are also present in the majority of the Skp1 homologs from protists, fungi, animals, and plants. The single Skp1 homologs from protists, fungi, and other vertebrates usually have the same critical residues as in the human Skp1. In *C. elegans*, *D. melanogaster*, and many plant species, when multiple Skp1 homologs exist, only the slowly evolving members have the same residues as or functionally very similar ones to those in the human Skp1. Our EST mining investigations and RT-PCR studies further indicated that these slowly evolving genes are usually expressed widely and/or at high levels (figs. 1 and 4), suggesting that they may be maintaining the most fundamental function(s) to interact with Cullin and F-box proteins.

In contrast, the rapidly evolving Skp1 homologs usually do not have EST information, suggesting that they are expressed at low levels, in specific tissues, or under specific conditions. Structurally, many of them differ from the human and fungal Skp1 at key positions. Some residues have been replaced by functionally similar amino acids, whereas others have been replaced by amino acids of quite different biochemical properties. Some of the moderately or rapidly evolving Skp1 homologs may have already lost the ability to interact with any Cullin and/or F-box protein and/or acquired new functions. The aforementioned Leu 105 of the human Skp1, for example, has been replaced by Met in *A. thaliana* Ask5, Ask6, and Ask10, by Asn in Ask7 and Ask8, and by Arg, Asp, and Lys in Ask3, Ask9, and Ask17, respectively (fig. 2; fig. 2 in Supplementary Material online). Our RT-PCR studies also indicated that the majority of *ASK* genes are expressed at low levels and/or in more specific fashion (fig. 4). The fact that *ASK6* was not detected through RT-PCR experiments further suggests that some of the moderately or rapidly evolving genes may have become pseudogenes.

The *ASK1* and *ASK2* genes seem to be the most important *SKP1* homologs in *A. thaliana*. Both of them are expressed widely and at high levels, and previous studies indicated that they may have overlapping functions. Genetic analysis indicates that *ASK1* is involved in regulating both vegetative and flower development and is essential for male meiosis (Yang et al. 1999; Zhao et al. 1999). In addition, *ASK1* affects responses to plant hormones auxin and jasmonic acid (Gray et al. 1999; Xu et al. 2002). The fact that the *ask1* null mutant has only mild defects in most of these developmental and physiological functions supports the hypothesis that *ASK1* and *ASK2* might be functionally redundant. Furthermore, Ask1 and Ask2 proteins have been shown to interact with the same proteins regulating flower development or hormonal signaling pathways (Gray et al. 1999; Samach et al. 1999; Xu et al. 2002), supporting the idea that they have similar functions. Based on the similarities in sequence, expression, and function, one may postulate that these two genes have a close relationship and/or may be a result of recent gene duplication. However, our phylogenetic analyses suggested that this is not the case (fig. 3*A*). The divergence of *ASK1*, *ASK2*, and *ASK4* occurred near the time of the eudicot radiation more than 125 MYA (Wikstrom, Savolainen, and Chase 2001). The very low d_N/d_S values for *ASK1, ASK2*, and *ASK4*, further indicate that these genes have critical functions and have been evolving under strong purifying selection over a minimum of 125 Myr. The maintenance of three highly conserved genes over such a long period suggests that these are not completely redundant, even though they still may share many common functions. The possibility of subfunctionalization (Lynch and Force 2000; Zhang 2003) is supported by the fact that the *ask1* null mutant is essential for male meiosis, a function that *ASK2* cannot fulfill. Even when the expression of the *ASK2* gene is elevated using another promoter, it could not fully rescue the *ask1* mutant defect in meiosis (Zhao et al. 2003). However, whether *ASK2* also has unique function(s) remains to be determined. Very little is known about the function of *ASK4*.

Figure 3 shows that sometimes the moderately or rapidly evolving Skp1 homologs from the same species or lineage can form one or several strongly supported clades, suggesting that some paralogs are the products of recent gene duplications. As indicated in other studies, duplications of individual genes, chromosomal segments, or entire genomes occurred so frequently in eukaryotic organisms that they have provided a primary source of material for the origin of evolutionary novelties (Wolfe and Shields 1997; The *Arabidopsis* Genome Initiative 2000; Knight 2002). In the *SKP1* gene family, recent gene duplications seem to have made great contributions to the increase of the gene numbers. In *A. thaliana*, for example, the tandemly arranged *ASK8*, *ASK7*, *ASK9*, and *ASK10* genes share amino acid sequence identities at levels higher than 80%, and all of them are expressed in the silique, although *ASK9* is also expressed elsewhere (fig. 4). Phylogenetically, these genes form a single clade, with bootstrap values higher than 90%. Therefore, *ASK7*, *ASK8*, *ASK9*, and *ASK10* may have originated from a single ancestral gene by recent gene duplications. Other groups of closely related genes that might be the result of recent duplications include *ASK11/ASK12* of *A. thaliana*, *SKR-8/SKR-9*, and *SKR-12/SKR-13* of *C. elegans*, and *OSK19/OSK22/OSK23* of rice, as supported by chromosome locations, sequence comparison, and phylogenetic analyses.

Conclusions and Perspectives

Our studies indicated that, in the animal and plant lineages, the multiple *SKP1* homologs from a single species have evolved at highly heterogeneous rates. The slowly evolving members are relatively conserved in sequence and expressed widely and/or at high levels, suggesting that they serve the most fundamental function(s), whereas the rapidly evolving members are structurally more diverse and usually expressed at low levels, in specific tissues, or possibly under specific conditions, suggesting that they may have lost the original function(s) and/or acquired new function(s). However, because of rate heterogeneity in evolution and long-branch attraction, it may not be possible to recover an accurate global phylogeny of the *SKP1* gene family. Variation in evolutionary constraint after gene duplications is likely

common, and rate heterogeneity among duplicate genes may confound phylogenetic analysis of many gene families. Our partition of the original data set according to evolutionary rates successfully addressed the problems in phylogenetic analysis and provided valuable information about the evolution of the *SKP1* gene family.

Our phylogenetic and d_N/d_S analyses suggested that the some *Arabidopsis ASK* genes (such as *ASK1*, *ASK2*, and *ASK4*) have evolved under strong purifying selection, whereas others are under weak purifying, neutral, or positive selection. As predicted (Lynch and Force 2000), closely related genes resulting from recent gene duplications usually have quite different d_N/d_S values, suggesting that they have evolved under different pressure and thus may have different function(s). Our results provide a phylogenetic context for additional functional studies to examine the existence of subfunctionalization and/or neofunctionalization during the evolution of *SKP1* genes.

Acknowledgments

We thank Helen Piontkivska for help with LINTREE analyses, Laura M. Zahn, Barbara Bliss, Liying Cui, Wei Hu and Dazhong Zhao for helpful discussions, and Michael Frohlich, Michael Purugganan, Jongmin Nam, Masatoshi Nei and two anonymous reviewers for critical reading and valuable comments on the manuscript. This work was supported by the National Science Foundation of the United States (DBI-0115684 and MCB-0092075), the National Natural Science Foundation of China (30100011 and 30130030), and funds from the Department of Biology and the Huck Institute for Life Sciences at the Pennsylvania State University.

Literature Cited

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.

Bai, C., P. Sen, K. Hofmann, L. Ma, M. Goebl, J. W. Harper, and S. J. Elledge. 1996. SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell **86**:263–274.

Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science **290**:972–977.

Callis, J., and R. D. Vierstra. 2000. Protein degradation in signaling. Curr. Opin. Plant Biol. **3**:381–386.

Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt. 1978. A model of evolutionary change in proteins: matrices for detecting distant relationships. Pp. 345–358 *in* M. O. Dayhoff, ed. Atlas of protein sequence and structure. National Biomedical Research Foundation, Washington, DC.

Farras, R., A. Ferrando, J. Jasik, T. Kleinow, L. Okresz, A. Tiburcio, K. Salchert, C. del Pozo, J. Schell, and C. Koncz. 2001. SKP1-SnRK protein kinase interactions mediate proteasomal binding of a plant SCF ubiquitin ligase. EMBO J. **20**:2742–2756.

Felsenstein, J. 2002. PHYLIP (phylogeny inference package), Version 3.6a3. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Gagne, J. M., B. P. Downes, S. H. Shiu, A. M. Durski, and R. D. Vierstra. 2002. The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **99**:11519–11524.

Ganoth, D., G. Bornstein, T. K. Ko, B. Larsen, M. Tyers, M. Pagano, and A. Hershko. 2001. The cell-cycle regulatory protein Cks1 is required for SCF^{Skp2}-mediated ubiquitinylation of p27. Nat. Cell Biol. **3**:321–324.

Goff, S. A., D. Ricke, T. H. Lan et al. (52 co-authors) 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science **296**:92–100.

Grassly, N. C., J. Adachi, and A. Rambaut. 1997. PSeq-Gen: an application for the Monte Carlo simulation of protein sequence evolution along phylogenetic trees. Comput. Appl. Biosci. **13**:559–560.

Gray, W. M., J. C. del Pozo, L. Walker, L. Hobbie, E. Risseeuw, T. Banks, W. L. Crosby, M. Yang, H. Ma, and M. Estelle. 1999. Identification of an SCF ubiquitin-ligase complex required for auxin response in *Arabidopsis thaliana*. Genes Dev. **13**:1678–1691.

Hatakeyama, S., M. Kitagawa, K. Nakayama et al. (11 co-authors) 1999. Ubiquitin-dependent degradation of IκBα is mediated by a ubiquitin ligase Skp1/Cul 1/F-box protein FWD1. Proc. Natl. Acad. Sci. USA **96**:3859–3863.

Hattori, K., S. Hatakeyama, M. Shirane, M. Matsumoto, and K. Nakayama. 1999. Molecular dissection of the interactions among IκBα, FWD1, and Skp1 required for ubiquitin-mediated proteolysis of IκBα. J. Biol. Chem. **274**:29641–29647.

Henikoff, S., and J. G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA **89**:10915–10919.

Hershko, A., and A. Ciechanover. 1998. The ubiquitin system. Annu. Rev. Biochem. **67**:425–479.

Huelsenbeck, J. P., D. M. Hillis, and R. Jones. 1996. Parametric bootstrapping in molecular phylogenetics: applications and performance. Pp. 19–45 *in* J. D. Ferris and S. R. Palumbi, eds. Molecular zoology: advances, strategies and protocols. Wiley-Liss, New York.

Ilyin, G., A. Serandour, C. Pigeon, M. Rialland, D. Glaise, and C. Guguen-Guillouzo. 2002. A new subfamily of structurally related human F-box proteins. Gene **296**:11.

Jackson, P. K., and A. G. Eldridge. 2002. The SCF ubiquitin ligase: an extended look. Mol. Cell **9**:923–925.

Kipreos, E. T., and M. Pagano. 2000. The F-box protein family. Genome Biol. 1:REVIEWS3002.

Knight, J. 2002. All genomes great and small. Nature **417**: 374–376.

Koepp, D. M., L. K. Schaefer, X. Ye, K. Keyomarsi, C. Chu, J. W. Harper, and S. J. Elledge. 2001. Phosphorylation-dependent ubiquitination of cyclin E by the SCF^{Fbw7} ubiquitin ligase. Science **294**:173–177.

Krek, W. 1998. Proteolysis and the G1-S transition: the SCF connection. Curr. Opin. Genet. Dev. **8**:36–42.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244–1245.

Lynch, M., and A. Force. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics **154**:459–473.

Nayak, S., F. E. Santiago, H. Jin, D. Lin, T. Schedl, and E. T. Kipreos. 2002. The *Caenorhabditis elegans* Skp1-related gene family: diverse functions in cell proliferation, morphogenesis, and meiosis. Curr. Biol. **12**:277–287.

Perkins, G., L. S. Drury, and J. F. Diffley. 2001. Separate SCF^{CDC4} recognition elements target Cdc6 for proteolysis in S phase and mitosis. EMBO J. **20**:4836–4845.

Samach, A., J. E. Klenz, S. E. Kohalmi, E. Risseeuw, G. W. Haughn, and W. L. Crosby. 1999. The *UNUSUAL FLORAL ORGANS* gene of *Arabidopsis thaliana* is an F-box protein

required for normal patterning and growth in the floral meristem. Plant J. **20**:433–445.

Schaffer, A. A., L. Aravind, T. L. Madden, S. Shavirin, J. L. Spouge, Y. I. Wolf, E. V. Koonin, and S. F. Altschul. 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res. **29**:2994–3005.

Schulman, B. A., A. C. Carrano, P. D. Jeffrey, Z. Bowen, E. R. Kinnucan, M. S. Finnin, S. J. Elledge, J. W. Harper, M. Pagano, and N. P. Pavletich. 2000. Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. Nature **408**:381–386.

Swofford, D. L. 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0 beta, Sinauer, Sunderland, Mass.

Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. **12**:823–833.

The *Arabidopsis* Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**:796–815.

Wikstrom, N., V. Savolainen, and M. W. Chase. 2001. Evolution of the angiosperms: calibrating the family tree. Proc. R. Soc. Lond. B Biol. Sci. **268**:2211–2220.

Willems, A. R., T. Goh, L. Taylor, I. Chernushevich, A. Shevchenko, and M. Tyers. 1999. SCF ubiquitin protein ligases and phosphorylation-dependent proteolysis. Philos. Trans. R Soc. Lond. B Biol. Sci. **354**:1533–1550.

Wolfe, K. H., and D. C. Shields. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature **387**:708–713.

Xu, L., F. Liu, E. Lechner, P. Genschik, W. L. Crosby, H. Ma, W. Peng, D. Huang, and D. Xie. 2002. The SCF^{COI1} ubiquitin-ligase complexes are required for jasmonate response in *Arabidopsis*. Plant Cell **14**:1919–1935.

Yamanaka, A., M. Yada, H. Imaki, M. Koga, Y. Ohshima, and K. Nakayama. 2002. Multiple Skp1-related proteins in *Caenorhabditis elegans*: diverse patterns of interaction with Cullins and F-box proteins. Curr. Biol. **12**:267–275.

Yang, M., Y. Hu, M. Lodhi, W. R. McCombie, and H. Ma. 1999. The *Arabidopsis SKP1-LIKE1* gene is essential for male meiosis and may control homologue separation. Proc. Natl. Acad. Sci. USA **96**:11416–11421.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**: 555–556.

Yu, J., S. Hu, J. Wang et al. (83 co-authors). 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science **296**:79–92.

Zhang, J. 2003. Evolution by gene duplication: an update. Trends Ecol. Evol. **18**:292–298.

Zhao, D., T. Han, E. Risseeuw, W. L. Crosby, and H. Ma. 2003. Conservation and divergence of *ASK1* and *ASK2* gene functions during male meiosis in *Arabidopsis thaliana*. Plant Mol. Biol. (in press).

Zhao, D., M. Yang, J. Solava, and H. Ma. 1999. The *ASK1* gene regulates development and interacts with the *UFO* gene to control floral organ identity in *Arabidopsis*. Dev. Genet. **25**:209–223.

Zhao, D., Q. Yu, M. Chen, and H. Ma. 2001. The *ASK1* gene regulates B function gene expression in cooperation with *UFO* and *LEAFY* in *Arabidopsis*. Development **128**:2735–2746.

Zheng, N., B. A. Schulman, L. Song et al. (14 co-authors). 2002. Structure of the Cul1-Rbx1-Skp1-F box^{Skp2} SCF ubiquitin ligase complex. Nature **416**:703–709.

William Jeffery, Associate Editor

Accepted August 12, 2003